



Guía docente 270107 - MD - Minería de Datos

Última modificación: 10/02/2025

Unidad responsable: Facultad de Informática de Barcelona
Unidad que imparte: 723 - CS - Departamento de Ciencias de la Computación.
715 - EIO - Departamento de Estadística e Investigación Operativa.

Titulación: GRADO EN INGENIERÍA INFORMÁTICA (Plan 2010). (Asignatura optativa).

Curso: 2024 **Créditos ECTS:** 6.0 **Idiomas:** Inglés

PROFESORADO

Profesorado responsable: MARIO MARTÍN MUÑOZ - CAROLINE LEONORE KÖNIG - CARINA GIBERT OLIVERAS

Otros: Primer quadrimestre:
XAVIER ANGERRI TORREDEFLOT - 11, 12
CARINA GIBERT OLIVERAS - 11, 12
MANUEL GIJON AGUDO - 11, 12
CAROLINE LEONORE KÖNIG - 11, 12

Segon quadrimestre:
XAVIER ANGERRI TORREDEFLOT - 11, 12
CARINA GIBERT OLIVERAS - 11, 12
MANUEL GIJON AGUDO - 11, 12
MARIO MARTÍN MUÑOZ - 11, 12
SERGI RAMIREZ MITJANS - 12

CAPACIDADES PREVIAS

Fundamentos de probabilidad y estadística. Programación básica en I

REQUISITOS

- Pre-requisito PE
- Pre-requisito PRO2

COMPETENCIAS DE LA TITULACIÓN A LAS QUE CONTRIBUYE LA ASIGNATURA

Específicas:

CSI2.2. Concebir, desplegar, organizar y gestionar sistemas y servicios informáticos, en contextos empresariales o institucionales, para mejorar sus procesos de negocio, responsabilizarse y liderar su puesta en marcha, y su mejora continua; y valorar su impacto económico y social.

CSI2.3. Demostrar conocimiento y capacidad de aplicación de los sistemas de extracción y de gestión del conocimiento.

CSI2.6. Demostrar conocimiento y capacidad de aplicación de los sistemas de ayuda a la toma de decisiones y de bussines intelligence.

Genéricas:

G3. TERCERA LENGUA: Conocer el idioma inglés con un nivel adecuado de forma oral y por escrito, y con consonancia con las necesidades que tendrán los graduados y graduadas en ingeniería informática. Capacidad de trabajar en un grupo multidisciplinar y en un entorno multilingüe, y de comunicar, tanto por escrito como de forma oral, conocimientos, procedimientos, resultados e ideas relacionadas con la profesión de ingeniero técnico en informática.

G9. RAZONAMIENTO: Capacidad de razonamiento crítico, lógico y matemático. Capacidad para resolver problemas dentro de su área de estudio. Capacidad de abstracción: capacidad de crear y utilizar modelos que reflejen situaciones reales. Capacidad de diseñar y realizar experimentos sencillos, y analizar e interpretar sus resultados. Capacidad de análisis, síntesis y evaluación.

METODOLOGÍAS DOCENTES

El aprendizaje se hará siguiendo la metodología de los casos, a partir del análisis de conjuntos de datos complejos provenientes de problemas reales. A partir de estos problemas se desarrollarán los conocimientos científicos necesarios en clase de teoría y su aplicación en las clases de laboratorio, de tal manera que la programación y/o integración de funciones de minería de datos reforzará la asimilación de los diferentes conceptos explicados. Para ello, se utilizará el entorno de programación libre R.

Las clases de laboratorio se dedicarán a resolver problemas relacionados con los conocimientos proporcionados en las clases de teoría y a la resolución por parte de los alumnos (de forma autónoma) de un problema similar. Este problema puede incluir la resolución de preguntas muy breves de carácter conceptual y será entregado para su evaluación. Por último, los alumnos deberán realizar dos prácticas completas, una de modelización estadística y otra para resolver problemas de tipo "científico", "transacciones" o "marketing" (uno a elegir). Esta última práctica será presentada de forma oral ante el conjunto de la clase.

OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

1. Conocer la tipología de los principales problemas de la Minería de Datos
2. Evaluación de la calidad de los datos y la necesidad de su preprocesado
3. Identificar las técnicas estadísticas y/o de aprendizaje automático más apropiadas al problema a resolver
5. Implementar algoritmos sencillos de aprendizaje
6. Evaluación de los resultados obtenidos
7. Presentación de los resultados en un entorno profesional para la toma de decisiones

HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTADO

Tipo	Horas	Porcentaje
Horas actividades dirigidas	6,0	4.00
Horas aprendizaje autónomo	84,0	56.00
Horas grupo pequeño	30,0	20.00
Horas grupo grande	30,0	20.00

Dedicación total: 150 h

CONTENIDOS

Introducción a la Minería de Datos.

Descripción:

Modelización estadística y tipos de problemas: análisis de datos binarios ("transacciones"), análisis de datos científicos y análisis de datos de empresas

Visualización y reducción de la dimensionalidad

Descripción:

Métodos de selección y extracción de variables. Visualización de datos multivariantes.



Clustering

Descripción:

Métodos de partición directa, jerárquicos y maximización de la esperanza

Métodos Predictivos

Descripción:

Regressió lineal múltiple i generalitzada. Regressió Logística. Xarxes Neuronals

Árboles de Decisión

Descripción:

Árboles de regresión y clasificación (CART).

Protocolos de validación y remuestreo de datos

Descripción:

Holdout, validación cruzada y bootstrap

Generación de reglas de asociación

Descripción:

Algoritmos A-priori y Eclat.

Métodos bayesianos

Descripción:

Teoría de la decisión bayesiana. Análisis Discriminante LDA, QDA. Naïve Bayes

Discriminación no paramétrica

Descripción:

Vecinos más cercanos

Métodos de contracción y selección de variables en regresión

Descripción:

Regressió lineal regularizada. Métodos LASSO y Elastic Net

Análisis formal de conceptos

Descripción:

Mètode formal para encontrar patrones en datos



Preprocesamiento de datos

Descripción:

a

Bagging i ensemble methods

Descripción:

Bagging i ensemble methods

ACTIVIDADES

Desarrollo del Tema 1

Objetivos específicos:

1

Dedicación: 2h

Grupo grande/Teoría: 2h

Repaso del lenguaje R

Dedicación: 6h

Grupo pequeño/Laboratorio: 6h

Desarrollo del tema 2

Objetivos específicos:

2

Dedicación: 16h

Aprendizaje autónomo: 8h

Grupo grande/Teoría: 4h

Grupo pequeño/Laboratorio: 4h

Desarrollo del tema 3

Objetivos específicos:

2

Dedicación: 9h

Aprendizaje autónomo: 4h

Grupo grande/Teoría: 3h

Grupo pequeño/Laboratorio: 2h



Desarrollo del Tema 4

Objetivos específicos:

2

Dedicación: 11h

Aprendizaje autónomo: 4h

Grupo grande/Teoría: 3h

Grupo pequeño/Laboratorio: 4h

Desarrollo del tema 5

Objetivos específicos:

2

Dedicación: 9h

Aprendizaje autónomo: 4h

Grupo grande/Teoría: 3h

Grupo pequeño/Laboratorio: 2h

Desarrollo del Tema 6

Objetivos específicos:

2

Dedicación: 7h

Aprendizaje autónomo: 4h

Grupo grande/Teoría: 3h

Desarrollo del Tema 7

Objetivos específicos:

2

Dedicación: 9h

Aprendizaje autónomo: 4h

Grupo grande/Teoría: 3h

Grupo pequeño/Laboratorio: 2h

Desarrollo del Tema 8

Objetivos específicos:

2

Dedicación: 11h

Aprendizaje autónomo: 4h

Grupo grande/Teoría: 3h

Grupo pequeño/Laboratorio: 4h



Desarrollo del Tema 9

Objetivos específicos:

2

Dedicación: 11h

Aprendizaje autónomo: 6h

Grupo grande/Teoría: 3h

Grupo pequeño/Laboratorio: 2h

Desarrollo del Tema 10

Objetivos específicos:

6

Competencias relacionadas:

G9. RAZONAMIENTO: Capacidad de razonamiento crítico, lógico y matemático. Capacidad para resolver problemas dentro de su área de estudio. Capacidad de abstracción: capacidad de crear y utilizar modelos que reflejen situaciones reales. Capacidad de diseñar y realizar experimentos sencillos, y analizar e interpretar sus resultados. Capacidad de análisis, síntesis y evaluación.

Dedicación: 13h

Aprendizaje autónomo: 6h

Grupo grande/Teoría: 3h

Grupo pequeño/Laboratorio: 4h

Practica 1

Objetivos específicos:

2, 3, 5, 6

Competencias relacionadas:

G9. RAZONAMIENTO: Capacidad de razonamiento crítico, lógico y matemático. Capacidad para resolver problemas dentro de su área de estudio. Capacidad de abstracción: capacidad de crear y utilizar modelos que reflejen situaciones reales. Capacidad de diseñar y realizar experimentos sencillos, y analizar e interpretar sus resultados. Capacidad de análisis, síntesis y evaluación.

Dedicación: 23h

Aprendizaje autónomo: 20h

Actividades dirigidas: 3h

Practica 2

Objetivos específicos:

3, 5, 6, 7

Competencias relacionadas:

G3. TERCERA LENGUA: Conocer el idioma inglés con un nivel adecuado de forma oral y por escrito, y con consonancia con las necesidades que tendrán los graduados y graduadas en ingeniería informática. Capacidad de trabajar en un grupo multidisciplinar y en un entorno multilingüe, y de comunicar, tanto por escrito como de forma oral, conocimientos, procedimientos, resultados e ideas relacionadas con la profesión de ingeniero técnico en informática.

G9. RAZONAMIENTO: Capacidad de razonamiento crítico, lógico y matemático. Capacidad para resolver problemas dentro de su área de estudio. Capacidad de abstracción: capacidad de crear y utilizar modelos que reflejen situaciones reales. Capacidad de diseñar y realizar experimentos sencillos, y analizar e interpretar sus resultados. Capacidad de análisis, síntesis y evaluación.

Dedicación: 23h

Aprendizaje autónomo: 20h

Actividades dirigidas: 3h

SISTEMA DE CALIFICACIÓN

La evaluación de la asignatura se realizará a partir de la nota obtenida en las ejercicios desarrollados durante las sesiones de laboratorio.

Por otra parte se realizarán dos prácticas, una realizada a partir de un problema de modelización estadística, y la otra práctica sobre uno de los otros tipos de problemas tratados en la asignatura. Por cada práctica el alumno presentará el correspondiente informe. Y por último, al finalizar el curso, el alumno deberá hacer la presentación oral de la segunda práctica.

El alumno deberá demostrar en estos informes y presentación haber adquirido los elementos de razonamiento e inglés de la asignatura. Estos se evaluarán mediante la correspondiente rúbrica.

La nota global de laboratorio es el promedio de las evaluaciones de los ejercicios desarrollados a partir de las sesiones de laboratorio.

La nota final se obtendrá así:

Labo = nota global de laboratorio

PR1 = nota de la primera práctica

PR2 = nota de la segunda práctica

Nota final = $0.2 * \text{Labo} + 0.4 * \text{PR1} + 0.4 * \text{PR2}$

En las dos prácticas, y respecto a su 40%, el 35% corresponde a la corrección técnica y el 5% corresponde a la competencia transversal 'razonamiento', de modo que se obtiene un peso global del 10% de esta competencia transversal en la nota final.

BIBLIOGRAFÍA

Básica:

- Hand, D.J. Construction and assessment of classification rules. Wiley, 1997. ISBN 978-0-471-96583-1.
- Hastie, T.; Tibshirani, R.; Friedman, J. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. Springer, 2009. ISBN 9780387848570.
- Hernández Orallo, J.; Ramírez Quintana, M.J.; Ferri Ramírez, C. Introducción a la minería de datos. Pearson, 2004. ISBN 9788420540917.
- Maindonald, J.H.; Braun, J. Data analysis and graphics using R: an example-based approach. 3rd ed. Cambridge University, 2010. ISBN 9780521762939.
- Duda, R.O.; Hart, P.E.; Stork, D.G. Pattern classification. 2nd ed. John Wiley & Sons, 2001. ISBN 0-471-05669-3.

Complementaria:

- Aluja Banet, T.; Morineau, A. Aprender de los datos: el análisis de componentes principales: una aproximación desde el Data Mining. EUB, 1999. ISBN 9788483120224.

RECURSOS

Enlace web:

- <http://www.cran.es.r-project.org>- <http://www.cs.waikako.ac.nz>- <http://www.kdnuggets.com/>